Wayback Machine

The **Wayback Machine** is a digital archive of the World Wide Web founded by the Internet Archive, an American nonprofit organization based in San Francisco, California. Created in 1996 and launched to the public in 2001, it allows users to go "back in time" to see how websites looked in the past. Its founders, Brewster Kahle and Bruce Gilliat, developed the Wayback Machine to provide "universal access to all knowledge" by preserving archived copies of defunct web pages.[2]

Launched on May 10, 1996, the Wayback Machine had saved more than 38.2 billion web pages by the end of 2009. As of November 2024, the Wayback Machine has archived more than 916 billion web pages and well over 100 petabytes of data.[3][4]

Wayback Machine

Type of site	Archive
Founded	May 10, 1996 (private) October 24, 2001 (public)
Area served	Worldwide (except China, India[a], and Bahrain)
Owner	Internet Archive
URL	web.archive.org (https://web.archive.org)
Commercial	No
Registration	Optional
Current status	Active
Written in	HTML, CSS, JavaScript, Java, Python

History

The Internet Archive began archiving cached web pages in 1996. One of the earliest known pages was archived on May 10, 1996, at 2:08 p.m. (UTC).[5]

Internet Archive founders Brewster Kahle and Bruce Gilliat launched the Wayback Machine in San Francisco, California,[6] in October 2001,[7][8] primarily to address the problem of web content vanishing whenever it gets changed or when a website is shut down.[9] The service enables users to see archived versions of web pages across time, which the archive calls a "three-dimensional index".[10] Kahle and Gilliat created the machine hoping to archive the entire Internet and provide "universal access to all knowledge".[11] The name "Wayback Machine" is a reference to a fictional time-traveling device in the animated cartoon *The Adventures of Rocky and Bullwinkle and Friends* from the 1960s.[12][13][14] In a segment of the cartoon entitled "Peabody's Improbable History", the characters Mister Peabody and Sherman use the "Wayback Machine" to travel back in time to witness and participate in famous historical events.[15]

From 1996 to 2001, the information was kept on digital tape, with Kahle occasionally allowing researchers and scientists to tap into the "clunky" database.[16] When the archive reached its fifth anniversary in 2001, it was unveiled and opened to the public in a ceremony at the University of California, Berkeley.[17] By the time the Wayback Machine launched, it already contained over 10 billion archived pages.[18] The data is stored on the Internet Archive's large cluster of Linux nodes.[11] It revisits and archives new versions of websites on occasion (see technical details below).[19] Sites can also be captured manually by entering a website's URL into the search box, provided that the website allows the Wayback Machine to "crawl" it and save the data.[20]

On October 30, 2020, the Wayback Machine began fact-checking content.[21] As of January 2022, domains of ad servers are disabled from capturing.[22]

In May 2021, for Internet Archive's 25th anniversary, the Wayback Machine introduced the "Wayforward Machine" which allows users to "travel to the Internet in 2046, where knowledge is under siege".[23][24]

Technical information

The Wayback Machine's software has been developed to "crawl" the Web and download all publicly accessible information and data files on webpages, the Gopher hierarchy, the Netnews (Usenet) bulletin board system, and downloadable software.[25] The information collected by these "crawlers" does not include all the information available on the Internet,

since much of the data is restricted by the publisher or stored in databases that are not accessible. To overcome inconsistencies in partially cached websites, Archive-It.org was developed in 2005 by the Internet Archive as a means of allowing institutions and content creators to voluntarily harvest and preserve collections of digital content, and create digital archives.[26]

Crawls are contributed from various sources, some imported from third parties and others generated internally by the Archive.[19] For example, crawls are contributed by the Sloan Foundation and Alexa, crawls run by Internet Archive on behalf of NARA and the Internet Memory Foundation, mirrors of Common Crawl.[19] The "Worldwide Web Crawls" have been running since 2010 and capture the global Web.[19][27] In September 2020, the Internet Archive announced a partnership with Cloudflare – an American content delivery network service provider – to automatically index websites served via its "Always Online" services.[28]

Documents and resources are stored with time stamp URLs such as `20241230173024`. Pages' individual resources such as images and style sheets and scripts, as well as outgoing hyperlinks, are linked to with the time stamp of the currently viewed page, so they are redirected automatically to their individual captures that are the closest in time.[29]

The frequency of snapshot captures varies per website.[19] Websites in the "Worldwide Web Crawls" are included in a "crawl list", with the site archived once per crawl.[19] A crawl can take months or even years to complete, depending on size.[19] For example, "Wide Crawl Number 13" started on January 9, 2015, and completed on July 11, 2016.[30] However, there may be multiple crawls ongoing at any one time, and a site might be included in more than one crawl list, so how often a site is crawled varies widely.[19]

A "Save Page Now" archiving feature was made available in October 2013,[31] accessible on the lower right of the Wayback Machine's main page.[32] Once a target URL is entered and saved, the web page will become part of the Wayback Machine.[31] Through the Internet address web.archive.org,[33] users can upload to the Wayback Machine a large variety of contents, including PDF and data compression file formats. The Wayback Machine creates a permanent local URL of the upload content, that is accessible in the web, even if not listed while searching in the https://archive.org official website.

Starting in October 2019, users were limited to 15 archival requests and retrievals per minute.[34]

Storage capacity and growth

As technology has developed over the years, the storage capacity of the Wayback Machine has grown. In 2003, after only two years of public access, the Wayback Machine was growing at a rate of 12 terabytes per month. The data is stored on PetaBox rack systems custom designed by Internet Archive staff. The first 100TB rack became fully operational in June 2004, although it soon became clear that they would need much more storage than that.[35][36]

The Internet Archive migrated its customized storage architecture to Sun Open Storage in 2009, and hosts a new data centre in a Sun Modular Datacenter on Sun Microsystems' California campus.[37] As of 2009, the Wayback Machine contained approximately three petabytes of data and was growing at a rate of 100 terabytes each month.[38]

A new, improved version of the Wayback Machine, with an updated interface and a fresher index of archived content, was made available for public testing in 2011, where captures appear in a calendar layout with circles whose width visualizes the number of crawls each day, but no marking of duplicates with asterisks or an advanced search page.[39][40] A top toolbar was added to facilitate navigating between captures. A bar chart visualizes the frequency of captures per month over the years.[41] Features like "Changes", "Summary", and a graphical site map were added subsequently.

In March that year, it was said on the Wayback Machine forum that "the Beta of the new Wayback Machine has a more complete and up-to-date index of all crawled materials into 2010, and will continue to be updated regularly. The index driving the classic Wayback Machine only has a little bit of material past 2008, and no further index updates are planned, as it will be phased out this year."[42] Also in 2011, the Internet Archive installed their sixth pair of PetaBox racks which increased the Wayback Machine's storage capacity by 700 terabytes.[43]

In January 2013, the company announced a milestone of 240 billion URLs.[44]

In October 2013, the company introduced the "Save a Page" feature, which allows any Internet user to archive the contents of a URL, and quickly generates a permanent link unlike the preceding *liveweb* feature.[45][46]

In December 2014, the Wayback Machine contained 435 billion web pages—almost nine petabytes of data, and was growing at about 20 terabytes a week.[18][47][48]

In July 2016, the Wayback Machine reportedly contained around 15 petabytes of data.[49] In October 2016, it was announced that the way web pages are counted would be changed, resulting in the decrease of the archived pages counts shown. Embedded objects such as pictures, videos, style sheets, JavaScripts are no longer counted as a "web page", whereas HTML, PDF, and plain text documents remain counted.[50]

In September 2018, the Wayback Machine contained over 25 petabytes of data.[51][52] As of December 2020, the Wayback Machine contained over 70 petabytes of data.[53]

Wayback Machine growth[54][55]

Wayback Machine by year	Pages archived	
2004	30,000,000,000	(0–100B: Light blue)
2005	40,000,000,000	
2008	85,000,000,000	
2012	150,000,000,000	(100B–450B: Yellow)
2013	373,000,000,000	
2014	400,000,000,000	
2015	452,000,000,000	(450B–600B: Orange)
2016	459,000,000,000	
2017	279,000,000,000	
2018	310,000,000,000	
2019	345,000,000,000	
2020	405,000,000,000	
2021	514,000,000,000	
2022	640,000,000,000	(600B–: Red)
2024	866,000,000,000	

Wayback Machine APIs

The Wayback Machine service offers three public APIs, SavePageNow, Availability, and CDX.[56] SavePageNow can be used to archive web pages. Availability API for checking the archive availability status for a web page,[57] checking whether an archive for the web page exists or not. CDX API is for complex querying, filtering, and analysis of captured data.[58][59]

Website exclusion policy

Historically, the Wayback Machine has respected the robots exclusion standard (robots.txt) in determining if a website would be crawled – or if already crawled, if its archives would be publicly viewable. Website owners had the option to opt out of Wayback Machine through the use of robots.txt. It applied robots.txt rules retroactively; if a site blocked the Internet Archive, any previously archived pages from the domain were immediately rendered unavailable as well. In addition, the Internet Archive stated that "Sometimes, a website owner will contact us directly and ask us to stop crawling or archiving a site. We comply with these requests."[60] In addition, the website says: "The Internet Archive is not interested in preserving or offering access to Web sites or other internet documents of persons who do not want their materials in the collection."[61][62]

On April 17, 2017, reports surfaced of sites that had gone defunct and became parked domains that were using robots.txt to exclude themselves from search engines, resulting in them being inadvertently excluded from the Wayback Machine.[63] Following this, the Internet Archive changed the policy to require an explicit exclusion request to remove sites from the Wayback Machine.[29]

The Oakland Archive Policy

Wayback's retroactive exclusion policy is based in part upon *Recommendations for Managing Removal Requests and Preserving Archival Integrity*, known as *The Oakland Archive Policy*, published by the School of Information Management and Systems at University of California, Berkeley in 2002, which gives a website owner the right to block access to the site's archives.[64] Wayback has complied with this policy to help avoid expensive litigation.[65]

The Wayback retroactive exclusion policy began to relax in 2017, when it stopped honoring robots on U.S. government and military web sites for both crawling and displaying web pages. As of April 2017, Wayback is ignoring robots.txt more broadly, not just for U.S. government websites.[66][67][68][69]

Uses

From its public launch in 2001, the Wayback Machine has been studied by scholars both for the ways it stores and collects data as well as for the actual pages contained in its archive. As of 2013, scholars had written about 350 articles on the Wayback Machine, mostly from the information technology, library science, and social science fields. Social science scholars have used the Wayback Machine to analyze how the development of websites from the mid-1990s to the present has affected the company's growth.[18]

When the Wayback Machine archives a page, it usually includes most of the hyperlinks, keeping those links active when they just as easily could have been broken by the Internet's instability. Researchers in India studied the effectiveness of the Wayback Machine's ability to save hyperlinks in online scholarly publications and found that it saved slightly more than half of them.[70]

"Journalists use the Wayback Machine to view dead websites, dated news reports, and changes to website contents. Its content has been used to hold politicians accountable and expose battlefield lies."[71] In 2014, an archived social media page of Igor Girkin, a separatist rebel leader in Ukraine, showed him boasting about his troops having shot down a suspected Ukrainian military airplane before it became known that the plane actually was a civilian Malaysian Airlines jet (Malaysia Airlines Flight 17), after which he deleted the post and blamed Ukraine's military for downing the plane.[71][72] In 2017, the March for Science originated from a discussion on Reddit that indicated someone had visited Archive.org and discovered that all references to climate change had been deleted from the White House website. In response, a user commented, "There needs to be a Scientists' March on Washington".[73][74][75]

The site is used heavily for verification, providing access to references and content creation by Wikipedia editors.[76] When new URLs are added to Wikipedia, the Internet Archive has been archiving them.[76]

In September 2020, a partnership was announced with Cloudflare to automatically archive websites served via its "Always Online" service, which will also allow it to direct users to its copy of the site if it cannot reach the original host.[28]

Limitations

In 2014, there was a six-month lag time between when a website was crawled and when it became available for viewing in the Wayback Machine.[77] As of 2024, the lag time is 3 to 10 hours.[29] The Wayback Machine offers only limited search facilities. Its "Site Search" feature allows users to find a site based on words describing the site, rather than words found on the web pages themselves.[78]

The Wayback Machine does not include every web page ever made due to the limitations of its web crawler. The Wayback Machine cannot completely archive web pages that contain interactive features such as Flash platforms and forms written in JavaScript and progressive web applications, because those functions require interaction with the host website. This means that, since approximately July 9, 2013, the Wayback Machine has been unable to display YouTube comments when saving

videos' watch pages, as, according to the Archive Team, comments are no longer "loaded within the page itself."[79] The Wayback Machine's web crawler has difficulty extracting anything not coded in HTML or one of its variants, which can often result in broken hyperlinks and missing images. Due to this, the web crawler cannot archive "orphan pages" that are not linked to by other pages.[78][80] The Wayback Machine's crawler only follows a predetermined number of hyperlinks based on a preset depth limit, so it cannot archive every hyperlink on every page.[27]

In legal evidence

Civil litigation

Netbula LLC v. Chordiant Software Inc.

In a 2009 case, *Netbula, LLC v. Chordiant Software Inc*., defendant Chordiant filed a motion to compel Netbula to disable the robots.txt file on its website that was causing the Wayback Machine to retroactively remove access to previous versions of pages it had archived from Netbula's site, pages that Chordiant believed would support its case.[81]

Netbula objected to the motion on the ground that defendants were asking to alter Netbula's website and that they should have subpoenaed Internet Archive for the pages directly.[82] An employee of Internet Archive filed a sworn statement supporting Chordiant's motion, however, stating that it could not produce the web pages by any other means "without considerable burden, expense and disruption to its operations."[81]

Magistrate Judge Howard Lloyd in the Northern District of California, San Jose Division, rejected Netbula's arguments and ordered them to disable the robots.txt blockage temporarily in order to allow Chordiant to retrieve the archived pages that they sought.[81]

Telewizja Polska USA, Inc. v. Echostar Satellite

In an October 2004 case, *Telewizja Polska USA, Inc. v. Echostar Satellite*, No. 02 C 3293, 65 Fed. R. Evid. Serv. 673 (N.D. Ill. October 15, 2004), a litigant attempted to use the Wayback Machine archives as a source of admissible evidence, perhaps for the first time. Telewizja Polska is the provider of TVP Polonia and EchoStar operates the Dish Network. Prior to the trial proceedings, EchoStar indicated that it intended to offer Wayback Machine snapshots as proof of the past content of Telewizja Polska's website. Telewizja Polska brought a motion *in limine* to suppress the snapshots on the grounds of hearsay and unauthenticated source, but Magistrate Judge Arlander Keys rejected Telewizja Polska's assertion of hearsay and denied TVP's motion *in limine* to exclude the evidence at trial.[83][84] At the trial, however, District Court Judge Ronald Guzman, the trial judge, overruled Magistrate Keys' findings, and held that neither the affidavit of the Internet Archive employee nor the underlying pages (i.e., the Telewizja Polska website) were admissible as evidence. Judge Guzman reasoned that the employee's affidavit contained both hearsay and inconclusive supporting statements, and the purported web page, printouts were not self-authenticating.[85][86]

Patent law

The United States Patent and Trademark Office and the European Patent Office will accept date stamps from the Internet Archive as evidence of when a given Web page was accessible to the public. These dates are used to determine if a Web page is available as prior art for instance in examining a patent application.[87]

Limitations of utility

There are technical limitations to archiving a website, and as a consequence, opposing parties in litigation can misuse the results provided by website archives. This problem can be exacerbated by the practice of submitting screenshots of web pages in complaints, answers, or expert witness reports when the underlying links are not exposed and therefore, can contain errors. For example, archives such as the Wayback Machine do not fill out forms and therefore, do not include the contents of non-RESTful e-commerce databases in their archives.[88]

Legal status

In Europe, the Wayback Machine could be interpreted as violating copyright laws. Only the content creator can decide where their content is published or duplicated so the Archive would have to delete pages from its system upon request of the creator.[89] The exclusion policies for the Wayback Machine may be found in the FAQ section of the site.[90]

Some cases have been brought against the Internet Archive specifically for its Wayback Machine archiving efforts.

Archived content legal issues

Scientology

In late 2002, the Internet Archive removed various sites that were critical of Scientology from the Wayback Machine.[91] An error message stated that this was in response to a "request by the site owner".[92] Later, it was clarified that lawyers from the Church of Scientology had demanded the removal and that the site owners did not want their material removed.[93]

Healthcare Advocates, Inc.

In 2003, Harding Earley Follmer & Frailey defended a client from a trademark dispute using the Archive's Wayback Machine. The attorneys were able to demonstrate that the claims made by the plaintiff were invalid, based on the content of their website from several years prior. The plaintiff, Healthcare Advocates, then amended their complaint to include the Internet Archive, accusing the organization of copyright infringement as well as violations of the DMCA and the Computer Fraud and Abuse Act. Healthcare Advocates claimed that, since they had installed a robots.txt file on their website, even if after the initial lawsuit was filed, the Archive should have removed all previous copies of the plaintiff website from the Wayback Machine, however, some material continued to be publicly visible on Wayback.[94] The lawsuit was settled out of court after Wayback fixed the problem.[95]

Suzanne Shell

Activist Suzanne Shell filed suit in December 2005, demanding Internet Archive pay her US$100,000 for archiving her website profane-justice.org between 1999 and 2004.[96][97] Internet Archive filed a declaratory judgment action in the United States District Court for the Northern District of California on January 20, 2006, seeking a judicial determination that Internet Archive did not violate Shell's copyright. Shell responded and brought a countersuit against Internet Archive for archiving her site, which she alleges is in violation of her terms of service.[98] On February 13, 2007, a judge for the United States District Court for the District of Colorado dismissed all counterclaims except breach of contract.[97] The Internet Archive did not move to dismiss the copyright infringement claims that Shell asserted arose out of its copying activities, which would also go forward.[99]

On April 25, 2007, Internet Archive and Suzanne Shell jointly announced the settlement of their lawsuit.[96] The Internet Archive said it "...has no interest in including materials in the Wayback Machine of persons who do not wish to have their Web content archived. We recognize that Ms. Shell has a valid and enforceable copyright in her Web site and we regret that the inclusion of her Web site in the Wayback Machine resulted in this litigation." Shell said, "I respect the historical value of Internet Archive's goal. I never intended to interfere with that goal nor cause it any harm."[100]

Daniel Davydiuk

Between 2013 and 2016, a pornographic actor named Daniel Davydiuk tried to remove archived images of himself from the Wayback Machine's archive, first by sending multiple DMCA requests to the archive, and then by appealing to the Federal Court of Canada.[101][102][103] The images were removed from the website in 2017.

FlexiSpy

In 2018, archives of stalkerware application FlexiSpy's website were removed from the Wayback Machine. The company claimed to have contacted the Internet Archive, presumably to remove the archives of its website.[104]

Censorship and other threats

Archive.org is blocked in China.[105][106][107] The Internet Archive was blocked in its entirety in Russia in 2015–16, ostensibly for hosting a Jihad outreach video.[71][108][109] Since 2016, the website has been back, available in its entirety, although in 2016 Russian commercial lobbyists were suing the Internet Archive to ban it on copyright grounds.[110]

In March 2015, it was published that security researchers became aware of the threat posed by the service's unintentional hosting of malicious binaries from archived sites.[111][112]

Alison Macrina, director of the Library Freedom Project, notes that "while librarians deeply value individual privacy, we also strongly oppose censorship".[71]

There is at least one case in which an article was removed from the archive shortly after it had been removed from its original website. A *Daily Beast* reporter had written an article that outed several gay Olympian athletes in 2016 after the reporter had made a fake profile posing as a gay man on a dating app. *The Daily Beast* removed the article after it was met with widespread furor; not long after, the Internet Archive soon did as well, but emphatically stated that they did so for no other reason than to protect the safety of the outed athletes.[71]

Other threats include natural disasters,[113] destruction (both remote and physical),[114] manipulation of the archive's contents, problematic copyright laws,[115] and surveillance of the site's users.[116]

Alexander Rose, executive director of the Long Now Foundation, suspects that in the long term of multiple generations "next to nothing" will survive in a useful way, stating, "If we have continuity in our technological civilization, I suspect a lot of the bare data will remain findable and searchable. But I suspect almost nothing of the format in which it was delivered will be recognizable" because sites "with deep back-ends of content-management systems like Drupal and Ruby and Django" are harder to archive.[117]

In 2016, in an article reflecting on the preservation of human knowledge, *The Atlantic* has commented that the Internet Archive, which describes itself to be built for the long-term,[118] "is working furiously to capture data before it disappears without any long-term infrastructure to speak of."[119]

In September 2024, the Internet Archive suffered a data breach that exposed 31 million records containing personal information, including email addresses and hashed passwords.[120] On October 9, 2024, the site went down due to a distributed denial-of-service attack.[121][122] On October 14, the site returned online, but it remained in read-only mode until November 4, during which time "Save Page Now" was disabled, replaced with a "Temporarily Unavailable" banner.[123]

See also

- Anna's Archive
- archive.today
- Heritrix
- Library Genesis
- Link rot
- List of Web archiving initiatives
- Time capsule

- Z-Library

Notes

a. Although it is formally blocked, enforcement is not consistent and depends on the region.[1]

References

1. Ong, Thuy (August 9, 2017). "Wayback Machine has been blocked in India" (https://www.theverge.com/2017/8/9/16117578/wayback-machine-blocked-india-internet-archive). *The Verge*.
2. Kahle, Brewster (November 23, 2005). "Universal Access to all Knowledge" (https://web.archive.org/web/20220814164546/https://archive.org/details/SDForumBK). *Internet Archive*. Archived from the original (https://archive.org/details/SDForumBK) on August 14, 2022. Retrieved June 5, 2022.
3. "Internet Archive: Wayback Machine" (https://web.archive.org/web/20230313021854/https://archive.org/web/). *web.archive.org*. Archived from the original (https://web.archive.org/) on March 13, 2023. The current number of archived pages can be seen at the archive's home page (https://web.archive.org/).
4. Kahle, Brewster. "A Message from Internet Archive Founder, Brewster Kahle" (https://archive.org/donate). *Internet Archive*. Retrieved January 10, 2024.
5. PepsiCo, Inc. (May 10, 1996). "PepsiCo Home Page" (https://web.archive.org/web/19960510140808/http://www.pepsico.com/). *Internet Archive/Wayback Machine*. Archived from the original (http://www.pepsico.com/) on May 10, 1996. Retrieved October 8, 2022.
6. "Wayback Machine General Information" (https://archive.today/20191205053755/https://help.archive.org/hc/en-us/articles/360004716091-Wayback-Machine-General-Information). *Internet Archive*. Archived from the original (https://help.archive.org/hc/en-us/articles/360004716091-Wayback-Machine-General-Information) on December 5, 2019. Retrieved March 2, 2021.
7. "WayBackMachine.org WHOIS, DNS, & Domain Info – DomainTools" (http://whois.domaintools.com/waybackmachine.org). *WHOIS*. Archived (https://web.archive.org/web/20200514175402/http://whois.domaintools.com/waybackmachine.org) from the original on May 14, 2020. Retrieved March 13, 2016.
8. "InternetArchive.org WHOIS, DNS, & Domain Info – DomainTools" (http://whois.domaintools.com/internetarchive.org). *WHOIS*. Archived (https://web.archive.org/web/20200512114745/http://whois.domaintools.com/internetarchive.org) from the original on May 12, 2020. Retrieved March 13, 2016.
9. Notess, Greg R. (March–April 2002). "The Wayback Machine: The Web's Archive". *Online*. **26**: 59–61. INIST 13517724 (https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=13517724).
10. "The Wayback Machine" (https://archive.org/about/faqs.php#The_Wayback_Machine), *Frequently Asked Questions*, archived (https://web.archive.org/web/20180918131957/https://archive.org/about/faqs.php#The_Wayback_Machine) from the original on September 18, 2018, retrieved September 18, 2018
11. "20,000 Hard Drives on a Mission" (https://blog.archive.org/2016/10/25/20000-hard-drives-on-a-mission/). *Internet Archive Blogs*. October 25, 2016. Archived (https://web.archive.org/web/20181020153727/https://blog.archive.org/2016/10/25/20000-hard-drives-on-a-mission/) from the original on October 20, 2018. Retrieved October 15, 2018.
12. Green, Heather (February 28, 2002). "A Library as Big as the World" (https://web.archive.org/web/20111220074306/http://www.businessweek.com/technology/content/feb2002/tc20020228_1080.htm). *BusinessWeek*. Archived from the original (http://www.businessweek.com/technology/content/feb2002/tc20020228_1080.htm) on December 20, 2011.
13. Tong, Judy (September 8, 2002). "Responsible Party – Brewster Kahle; A Library Of the Web, On the Web" (https://www.nytimes.com/2002/09/08/business/responsible-party-brewster-kahle-a-library-of-the-web-on-the-web.html). *The New York Times*. Archived (https://web.archive.org/web/20110220181725/http://www.nytimes.com/2002/09/08/business/responsible-party-brewster-kahle-a-library-of-the-web-on-the-web.html) from the original on February 20, 2011. Retrieved August 15, 2011.
14. Keith Scott (2000). *The Moose that Roared: The Story of Jay Ward, Bill Scott, a Flying Squirrel, and a Talking Moose (https://books.google.com/books?id=qdwcbJxM-HsC)*. St. Martin's Press. ISBN 0-312-19922-8
15. Markstein, Don. "Toonopedia: "Peabody's Improbable History" " (http://www.toonopedia.com/peabody.htm). Toonpedia.com. Retrieved November 7, 2024.
16. Cook, John (November 1, 2001). "Web site takes you way back in Internet history" (http://www.seattlepi.com/news/article/Web-site-takes-you-way-back-in-Internet-history-1070534.php). *Seattle Post-Intelligencer*. Archived (https://web.archive.org/web/20140812205508/http://www.seattlepi.com/news/article/Web-site-takes-you-way-back-in-Internet-history-1070534.php) from the original on August 12, 2014. Retrieved August 15, 2011.

17. Mayfield, Kendra (October 28, 2001). "Wayback Goes Way Back on Web" (https://www.wired.com/2001/10/wayback-goes-way-back-on-web/). *Wired*. Archived (https://web.archive.org/web/20171016225914/https://www.wired.com/2001/10/wayback-goes-way-back-on-web/) from the original on October 16, 2017. Retrieved October 16, 2017.
18. Arora, Sanjay K.; Li, Yin; Youtie, Jan; Shapira, Philip (May 5, 2015). "Using the wayback machine to mine websites in the social sciences: A methodological resource" (https://doi.org/10.1002%2Fasi.23503). *Journal of the Association for Information Science and Technology*. **67** (8): 1904–1915. doi:10.1002/asi.23503 (https://doi.org/10.1002%2Fasi.23503). hdl:10.1002/asi.23503 (https://hdl.handle.net/10.1002%2Fasi.23503). ISSN 2330-1635 (https://search.worldcat.org/issn/2330-1635).
19. Leetaru, Kalev (January 28, 2016). "The Internet Archive Turns 20: A Behind the Scenes Look at Archiving the Web" (https://www.forbes.com/sites/kalevleetaru/2016/01/18/the-internet-archive-turns-20-a-behind-the-scenes-look-at-archiving-the-web/#222f2e5682e0). *Forbes*. Archived (https://web.archive.org/web/20171016230439/https://www.forbes.com/sites/kalevleetaru/2016/01/18/the-internet-archive-turns-20-a-behind-the-scenes-look-at-archiving-the-web/#222f2e5682e0) from the original on October 16, 2017. Retrieved October 16, 2017.
20. "Internet Archive: Wayback Machine" (https://archive.org/web/). *Internet Archive*. Archived (https://web.archive.org/web/20140103004344/https://archive.org/web/) from the original on January 3, 2014. Retrieved October 15, 2018.
21. Graham, Mark (October 30, 2020). "Fact Checks and Context for Wayback Machine Pages" (http://blog.archive.org/2020/10/30/fact-checks-and-context-for-wayback-machine-pages/). *Internet Archive Blogs*. Retrieved January 17, 2021.
22. Attempts to 'save page now' domains such as tpc.googlesyndication.com or s0.2mdn.net or atdmt.com or adbrite.com result in "This URL is in our block list and cannot be captured."
23. "Internet Archive 25th Anniversary – Universal Access to All Knowledge" (https://anniversary.archive.org/). Retrieved January 13, 2022.
24. "Wayforward Machine • Visit the future of the internet" (https://wayforward.archive.org/). *Way Forward Machine*. Retrieved January 13, 2022.
25. Kahle, Brewster. "Archiving the Internet" (http://www.uibk.ac.at/voeb/texte/kahle.html). Scientific American – March 1997 Issue. Archived (https://web.archive.org/web/20120403042627/http://www.uibk.ac.at/voeb/texte/kahle.html) from the original on April 3, 2012. Retrieved August 19, 2011.
26. Kaplan, Jeff (October 27, 2014). "Archive-It: Crawling the Web Together" (https://blog.archive.org/2014/10/27/archive-it-crawling-the-web-together/). *Internet Archive Blogs*. Archived (https://web.archive.org/web/20171012212827/http://blog.archive.org/2014/10/27/archive-it-crawling-the-web-together/) from the original on October 12, 2017. Retrieved October 16, 2017.
27. "Worldwide Web Crawls" (https://archive.org/details/widecrawl&tab=about). Internet Archive. Archived (https://web.archive.org/web/20171019222740/https://archive.org/details/widecrawl%26tab%3Dabout) from the original on October 19, 2017. Retrieved October 16, 2017.
28. Graham, Mark (September 17, 2020). "Cloudflare and the Wayback Machine, joining forces for a more reliable Web" (http://blog.archive.org/2020/09/17/internet-archive-partners-with-cloudflare-to-help-make-the-web-more-useful-and-reliable/). *Internet Archive Blogs*. Retrieved September 17, 2020.
29. "Using The Wayback Machine" (https://help.archive.org/help/using-the-wayback-machine/). *Internet Archive*. Retrieved September 25, 2024.
30. "Wide Crawl Number 13" (https://archive.org/details/wide00013?&sort=-publicdate&page=3). Internet Archive. Archived (https://web.archive.org/web/20171019223332/https://archive.org/details/wide00013?&sort=-publicdate&page=3) from the original on October 19, 2017. Retrieved October 16, 2017.
31. Rossi, Alexis (October 25, 2013). "Fixing Broken Links on the Internet" (https://blog.archive.org/2013/10/25/fixing-broken-links/). Internet Archive. Archived (https://web.archive.org/web/20141107193437/http://blog.archive.org/2013/10/25/fixing-broken-links/) from the original on November 7, 2014. Retrieved December 29, 2013.
32. "Wayback Machine main page" (https://archive.org/web/). Internet Archive. Archived (https://web.archive.org/web/20140103004344/https://archive.org/web/) from the original on January 3, 2014. Retrieved December 30, 2013.
33. "Web.archive.org directory" (https://web.archive.org). Archived (https://web.archive.org/web/20120103040016/https://web.archive.org/) from the original on January 3, 2012. Retrieved March 2, 2014.
34. "Too Many Requests" (https://archive.org/details/toomanyrequests_20191110). Internet Archive. November 10, 2019. Retrieved November 27, 2021.
35. "Petabox" (https://archive.org/web/petabox.php). *Internet Archive*. Retrieved October 25, 2018.
36. Kanellos, Michael (July 29, 2005). "Big storage on the cheap" (https://web.archive.org/web/20070403030705/http://news.zdnet.com/2100-9584_22-5808754.html). CNET News. Archived from the original (http://news.zdnet.com/2100-9584_22-5808754.html) on April 3, 2007. Retrieved July 29, 2007.

37. "Internet Archive and Sun Microsystems Create Living History of the Internet" (https://web.archive.org/web/20090326200212/http://www.sun.com/aboutsun/pr/2009-03/sunflash.20090325.1.xml). Sun Microsystems. March 25, 2009. Archived from the original (http://www.sun.com/aboutsun/pr/2009-03/sunflash.20090325.1.xml) on March 26, 2009. Retrieved March 27, 2009.
38. Mearian, Lucas (March 19, 2009). "Internet Archive to unveil massive Wayback Machine data center" (https://web.archive.org/web/20090323093002/http://www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyName=hardware&articleId=9130081&taxonomyId=12&intsrc=kc_top). Computerworld. Archived from the original (http://www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyName=hardware&articleId=9130081&taxonomyId=12&intsrc=kc_top) on March 23, 2009. Retrieved March 22, 2009.
39. gojomo (January 24, 2011). "Updated Wayback Machine in Beta Testing" (https://web.archive.org/web/20110823040310/http://iawebarchiving.wordpress.com/2011/01/24/updated-wayback-machine-in-beta-testing/). Archived from the original (http://iawebarchiving.wordpress.com/2011/01/24/updated-wayback-machine-in-beta-testing/) on August 23, 2011. Retrieved August 19, 2011.
40. "Advanced Search" (https://web.archive.org/web/20100131104918/https://web.archive.org/collections/web/advanced.html). *Wayback Machine*. Archived from the original (https://web.archive.org/collections/web/advanced.html) on January 31, 2010. Retrieved April 3, 2022.
41. "What's the difference between the classic Wayback Machine and the new Beta version?" (https://web.archive.org/web/20101225023945/http://faq.waybackmachine.org/whats-the-difference-between-the-classic-wayback-machine-and-the-new-beta-version). Archived from the original (http://faq.waybackmachine.org/whats-the-difference-between-the-classic-wayback-machine-and-the-new-beta-version) on December 25, 2010. Retrieved November 17, 2021.
42. "Beta Wayback Machine, in forum" (https://www.archive.org/post/350738/updated-wayback-machine-in-beta-testing). Archived (https://web.archive.org/web/20140417082107/https://archive.org/post/350738/updated-wayback-machine-in-beta-testing) from the original on April 17, 2014. Retrieved April 16, 2014.
43. "Internet Archive Forums: 6th pair of racks go into service: over 2PB of data space used" (https://archive.org/post/353721/6th-pair-of-racks-go-into-service-over-2pb-of-data-space-used). *Internet Archive*. Archived (https://web.archive.org/web/20161024144627/https://archive.org/post/353721/6th-pair-of-racks-go-into-service-over-2pb-of-data-space-used) from the original on October 24, 2016. Retrieved October 25, 2018.
44. "Wayback Machine: Now with 240,000,000,000 URLs | Internet Archive Blogs" (http://blog.archive.org/2013/01/09/updated-wayback/). January 9, 2013. Archived (https://web.archive.org/web/20140414221120/http://blog.archive.org/2013/01/09/updated-wayback/) from the original on April 14, 2014. Retrieved April 16, 2014.
45. Rossi, Alexis (October 25, 2013). "Fixing Broken Links on the Internet" (https://blog.archive.org/2013/10/25/fixing-broken-links/). *Internet Archive*. San Francisco, CA, US: Collections Team, the Internet Archive. Archived (https://web.archive.org/web/20141107193437/http://blog.archive.org/2013/10/25/fixing-broken-links/) from the original on November 7, 2014. Retrieved March 25, 2015. "We have added the ability to archive a page instantly and get back a permanent URL for that page in the Wayback Machine. This service allows anyone – wikipedia editors, scholars, legal professionals, students, or home cooks like me – to create a stable URL to cite, share or bookmark any information they want to still have access to in the future."
46. Baron, Alexander (October 23, 2013). "The new Internet Archive Wayback Machine now online" (https://web.archive.org/web/20201119071411/http://www.digitaljournal.com/article/360776). *Digital Journal*. Archived from the original (http://www.digitaljournal.com/article/360776) on November 19, 2020. Retrieved November 19, 2020.
47. "Internet Archive Frequently Asked Questions" (https://archive.org/about/faqs.php). Archived (https://web.archive.org/web/20091021003552/https://archive.org/about/faqs.php) from the original on October 21, 2009. Retrieved January 17, 2015.
48. "Internet Archive Frequently Asked Questions" (https://web.archive.org/web/20141218203115/https://archive.org/about/faqs.php). December 18, 2014. Archived from the original (https://archive.org/about/faqs.php) on December 18, 2014. Retrieved December 13, 2018.
49. "Can the manipulation of big data change the way the world thinks?" (http://www.thenational.ae/opinion/comment/can-the-manipulation-of-big-data-change-the-way-the-world-thinks). *The National*. Archived (https://web.archive.org/web/20170112060354/http://www.thenational.ae/opinion/comment/can-the-manipulation-of-big-data-change-the-way-the-world-thinks) from the original on January 12, 2017. Retrieved May 14, 2017.
50. Goel, Vinay (October 23, 2016). "Defining Web pages, Web sites and Web captures" (https://blog.archive.org/2016/10/23/defining-web-pages-web-sites-and-web-captures/). Internet Archive. Archived (https://web.archive.org/web/20181209195730/https://blog.archive.org/2016/10/23/defining-web-pages-web-sites-and-web-captures/) from the original on December 9, 2018. Retrieved December 9, 2018.
51. Crockett, Zachary (September 28, 2018). "Inside Wayback Machine, the internet's time capsule" (https://thehustle.co/inside-wayback-machine-internet-archive). *The Hustle*. Archived (https://web.archive.org/web/20181002145800/https://thehustle.co/inside-wayback-machine-internet-archive) from the original on October 2, 2018. Retrieved October 26, 2018.

52. Heffernan, Virginia (September 18, 2018). "Things Break and Decay on the Internet—That's a Good Thing" (https://www.wired.com/story/wired25-virginia-heffernan-internet-archive-wayback-machine/). *WIRED*. Archived (https://web.archive.org/web/20180925130510/https://www.wired.com/story/wired25-virginia-heffernan-internet-archive-wayback-machine/) from the original on September 25, 2018. Retrieved October 26, 2018.
53. "Donate to the Internet Archive: Digital Library of Free & Borrowable Books, Movies, Music & Wayback Machine @internetarchive" (https://blog.adafruit.com/2020/12/01/donate-to-the-internet-archive-digital-library-of-free-borrowable-books-movies-music-wayback-machine-internetarchive/). adafruit. December 1, 2020. Archived (https://web.archive.org/web/20201202065323/https://blog.adafruit.com/2020/12/01/donate-to-the-internet-archive-digital-library-of-free-borrowable-books-movies-music-wayback-machine-internetarchive/) from the original on December 2, 2020. Retrieved December 2, 2020.
54. michelle (May 9, 2014). "Wayback Machine Hits 400,000,000,000!" (https://blog.archive.org/2014/05/09/wayback-machine-hits-400000000000). Internet Archive. Archived (https://web.archive.org/web/20140826191225/http://blog.archive.org/2014/05/09/wayback-machine-hits-400000000000/) from the original on August 26, 2014. Retrieved March 25, 2015.
55. "Internet Archive" (https://web.archive.org/web/20201231000610/https://archive.org/). Internet Archive. Archived from the original (https://www.archive.org/) on December 31, 2020. Retrieved March 8, 2021.
56. "Wayback Machine APIs" (https://archive.org/help/wayback_api.php). *Internet Archive*.
57. waybackpy (https://github.com/akamhy/waybackpy) on GitHub
58. "Developers" (https://blog.archive.org/developers/). *Internet Archive Blogs*. August 22, 2014. Archived (https://web.archive.org/web/20240212160820/https://blog.archive.org/developers/) from the original on February 12, 2024.
59. Gonzalez, John (December 13, 2018). "Documentation for Public APIs at the Internet Archive" (http://blog.archive.org/2018/12/13/documentation-for-public-apis-at-the-internet-archive/). *Internet Archive Blogs*. Archived (https://web.archive.org/web/20240113211453/https://blog.archive.org/2018/12/13/documentation-for-public-apis-at-the-internet-archive/) from the original on January 13, 2024.
60. "FAQs – Some sites are not available because of Robots.txt or other exclusions. What does that mean?" (https://web.archive.org/web/20110415130934/https://web.archive.org/collections/web/faqs.html#exclusions). *Internet Archive Wayback Machine*. Archived from the original (https://web.archive.org/collections/web/faqs.html#exclusions) on April 15, 2011.
61. "Frequently Asked Questions" (https://web.archive.org/web/20140417122600/https://archive.org/about/faqs.php). *Internet Archive*. Archived from the original (https://www.archive.org/about/faqs.php#2) on April 17, 2014.
62. Cox, Joseph (May 22, 2018). "The Wayback Machine Is Deleting Evidence of Malware Sold to Stalkers" (https://motherboard.vice.com/en_us/article/nekzzq/wayback-machine-deleting-evidence-flexispy). *Vice*. Archived (https://archive.today/20180522192132/https://motherboard.vice.com/en_us/article/nekzzq/wayback-machine-deleting-evidence-flexispy) from the original on May 22, 2018. Retrieved May 23, 2018.
63. "Robots.txt meant for search engines don't work well for web archives" (https://blog.archive.org/2017/04/17/robots-txt-meant-for-search-engines-dont-work-well-for-web-archives/). *Internet Archive*. April 17, 2017. Retrieved June 29, 2019.
64. "Recommendations for Managing Removal Requests And Preserving Archival Integrity" (https://web.archive.org/web/20030502165937/http://sims.berkeley.edu/research/conferences/aps/removal-policy.html). University of California. December 14, 2002. Archived from the original (http://www2.sims.berkeley.edu/research/conferences/aps/removal-policy.html) on May 2, 2003. Retrieved October 20, 2024.
65. "Retroactive robots.txt removal of past crawls AKA Oakland Archive Policy" (https://archive.org/post/1019415/retroactive-robotstxt-removal-of-past-crawls-aka-oakland-archive-policy). Internet Archive. July 7, 2014. Archived (https://web.archive.org/web/20171010124036/https://archive.org/post/1019415/retroactive-robotstxt-removal-of-past-crawls-aka-oakland-archive-policy) from the original on October 10, 2017. Retrieved September 14, 2017.
66. Graham, Mark (April 17, 2017). "Robots.txt meant for search engines don't work well for web archives" (http://blog.archive.org/2017/04/17/robots-txt-meant-for-search-engines-dont-work-well-for-web-archives/). *Internet Archive Blogs*. Archived (https://web.archive.org/web/20170417131508/http://blog.archive.org/2017/04/17/robots-txt-meant-for-search-engines-dont-work-well-for-web-archives/) from the original on April 17, 2017. Retrieved April 16, 2017.
67. "Archivierung des Internets: Internet Archive ignoriert künftig robots.txt" (https://www.heise.de/newsticker/meldung/Archivierung-des-Internets-Internet-Archive-ignoriert-kuenftig-robots-txt-3693558.html) (in German). heise online. April 25, 2017. Archived (https://web.archive.org/web/20170427035659/https://www.heise.de/newsticker/meldung/Archivierung-des-Internets-Internet-Archive-ignoriert-kuenftig-robots-txt-3693558.html) from the original on April 27, 2017. Retrieved May 14, 2017.

68. "Suchmaschinen: Internet Archive will künftig Robots.txt-Einträge ignorieren – Golem.de" (https://www.golem.de/news/suchmaschinen-internet-archive-will-kuenftig-robots-txt-eintraege-ignorieren-1704-127446.html) (in German). Archived (https://web.archive.org/web/20170619210648/https://www.golem.de/news/suchmaschinen-internet-archive-will-kuenftig-robots-txt-eintraege-ignorieren-1704-127446.html) from the original on June 19, 2017. Retrieved May 14, 2017.
69. "Internet Archive will ignore robots.txt files to keep historical record accurate" (https://www.digitaltrends.com/computing/internet-archive-robots-txt/). *Digital Trends*. April 24, 2017. Archived (https://web.archive.org/web/20170516130029/https://www.digitaltrends.com/computing/internet-archive-robots-txt/) from the original on May 16, 2017. Retrieved May 14, 2017.
70. Sampath Kumar, B.T.; Prithviraj, K.R. (October 21, 2014). "Bringing life to dead: Role of Wayback Machine in retrieving vanished URLs". *Journal of Information Science*. **41** (1): 71–81. doi:10.1177/0165551514552752 (https://doi.org/10.1177%2F0165551514552752). ISSN 0165-5515 (https://search.worldcat.org/issn/0165-5515). S2CID 28320982 (https://api.semanticscholar.org/CorpusID:28320982).
71. Nelson, Steven (August 17, 2016). "Wayback Machine Won't Censor Archive for Taste, Director Says After Olympics Article Scrubbed" (https://www.usnews.com/news/articles/2016-08-17/wayback-machine-wont-censor-archive-for-taste-director-says-after-olympics-article-scrubbed). *U.S. News & World Report*. Archived (https://web.archive.org/web/20170106151933/http://www.usnews.com/news/articles/2016-08-17/wayback-machine-wont-censor-archive-for-taste-director-says-after-olympics-article-scrubbed) from the original on January 6, 2017. Retrieved May 14, 2017.
72. Lepore, Jill (January 26, 2015). "What the Web Said Yesterday" (https://www.newyorker.com/magazine/2015/01/26/cobweb). *The New Yorker*. Archived (https://web.archive.org/web/20150125141230/http://www.newyorker.com/magazine/2015/01/26/cobweb) from the original on January 25, 2015. Retrieved May 14, 2017.
73. Guarino, Ben (April 21, 2017). "The March for Science began with this person's 'throwaway line' on Reddit" (https://www.washingtonpost.com/news/speaking-of-science/wp/2017/04/21/the-march-for-science-began-with-this-persons-throwaway-line-on-reddit/). *Washington Post*. Archived (https://web.archive.org/web/20170423081417/https://www.washingtonpost.com/news/speaking-of-science/wp/2017/04/21/the-march-for-science-began-with-this-persons-throwaway-line-on-reddit/) from the original on April 23, 2017. Retrieved April 23, 2017.
74. Kaplan, Sarah (January 25, 2017). "Are scientists going to march on Washington?" (https://www.washingtonpost.com/news/speaking-of-science/wp/2017/01/24/are-scientists-going-to-march-on-washington/). *The Washington Post*. Archived (https://web.archive.org/web/20170131152535/https://www.washingtonpost.com/news/speaking-of-science/wp/2017/01/24/are-scientists-going-to-march-on-washington/) from the original on January 31, 2017. Retrieved January 31, 2017.
75. Foley, Katherine Ellen (April 22, 2017). "The global March for Science started with a single Reddit thread" (https://qz.com/965485/the-global-march-for-science-started-with-a-single-reddit-thread/). *Quartz*. Archived (https://web.archive.org/web/20170424004314/https://qz.com/965485/the-global-march-for-science-started-with-a-single-reddit-thread/) from the original on April 24, 2017. Retrieved April 23, 2017.
76. Graham, Mark (October 1, 2018). "More than 9 million broken links on Wikipedia are now rescued" (http://blog.archive.org/2018/10/01/more-than-9-million-broken-links-on-wikipedia-are-now-rescued/). *Internet Archive Blogs*. Archived (https://archive.today/20230408194542/http://blog.archive.org/2018/10/01/more-than-9-million-broken-links-on-wikipedia-are-now-rescued/) from the original on April 8, 2023.
77. "Internet Archive Frequently Asked Questions" (https://web.archive.org/web/20140402223358/https://archive.org/about/faqs.php). *Internet Archive*. April 2, 2014. Archived from the original (https://archive.org/about/faqs.php) on April 2, 2014. Retrieved November 23, 2018.
78. Bates, Mary Ellen (2002). "The Wayback Machine". *Online*. **26**: 80.
79. "YouTube – Archiveteam" (https://www.archiveteam.org/index.php?title=YouTube#Comment_loading). *archiveteam.org*. Archived (https://web.archive.org/web/20200805184742/https://www.archiveteam.org/index.php?title=YouTube#Comment_loading) from the original on August 5, 2020. Retrieved August 6, 2020.
80. "Internet Archive Frequently Asked Questions" (https://archive.org/about/faqs.php). *Internet Archive*. Archived (https://web.archive.org/web/20130420213122/https://archive.org/about/faqs.php) from the original on April 20, 2013. Retrieved October 18, 2018.
81. Lloyd, Howard (October 2009). "Order to Disable Robots.txt" (https://web.archive.org/web/20190808173832/http://www.american-justice.org/upload/page/123/69/docket-187-order-on-IA-motion.pdf) (PDF). Archived from the original (http://www.american-justice.org/upload/page/123/69/docket-187-order-on-IA-motion.pdf) (PDF) on August 8, 2019. Retrieved October 15, 2009.
82. Cortes, Antonio (October 2009). "Motion Opposing Removal of Robots.txt" (https://web.archive.org/web/20101027050350/http://www.american-justice.org/index.cgi/Page/116/OPPOSITION-TO-MOTION-TO-COMPEL-REMOVAL-OF-ROBOT-TXT-FILE-FROM-WEBSITE). Archived from the original (http://www.american-justice.org/index.cgi/Page/116/OPPOSITION-TO-MOTION-TO-COMPEL-REMOVAL-OF-ROBOT-TXT-FILE-FROM-WEBSITE/) on October 27, 2010. Retrieved October 15, 2009.

83. Gelman, Lauren (November 17, 2004). "Internet Archive's Web Page Snapshots Held Admissible as Evidence" (https://web.archive.org/web/20110430095339/http://cyberlaw.stanford.edu/packets002728.shtml). *Packets*. **2** (3). Archived from the original (http://cyberlaw.stanford.edu/packets002728.shtml) on April 30, 2011. Retrieved January 4, 2007.
84. Howell, Beryl A. (February 2006). "Proving Web History: How to use the Internet Archive" (https://web.archive.org/web/20100705043226/http://www.strozfriedberg.com/files/Publication/fee98a34-d739-478b-a7db-6af37b757714/Presentation/PublicationAttachment/aae88469-9835-4fe4-ae5f-38637924314f/BAHPROVINGWEBHISTORY.pdf) (PDF). *Journal of Internet Law*: 3–9. Archived from the original (http://www.strozfriedberg.com/files/Publication/fee98a34-d739-478b-a7db-6af37b757714/Presentation/PublicationAttachment/aae88469-9835-4fe4-ae5f-38637924314f/BAHPROVINGWEBHISTORY.pdf) (PDF) on July 5, 2010. Retrieved August 6, 2008.
85. "Looking For Evidence in Virtual Places Admissibility of Internet Evidence" (https://www.netforlawyers.com/page/looking-evidence-virtual-places-admissibility-internet-evidence). Archived (https://web.archive.org/web/20190701055139/https://www.netforlawyers.com/page/looking-evidence-virtual-places-admissibility-internet-evidence) from the original on July 1, 2019. Retrieved June 14, 2020.
86. Levitt, Carole A.; Rosch, Mark E. (2010). *Find Info Like a Pro: Mining the Internet's Publicly Available Resources for Investigative Research, Tom 1* (https://books.google.com/books?id=SUErZdbvcOkC). American Bar Association. pp. 194–196. ISBN 978-1-60442-890-2. Archived (https://web.archive.org/web/20201218143732/https://books.google.com/books?id=SUErZdbvcOkC) from the original on December 18, 2020. Retrieved June 14, 2020.
87. Coggins, Wynn W. (Fall 2002). "Prior Art in the Field of Business Method Patents – When is an Electronic Document a Printed Publication for Prior Art Purposes?" (https://web.archive.org/web/20120921083344/http://www.uspto.gov/patents/resources/methods/aiplafall02paper.jsp). *USPTO*. Archived from the original (http://www.uspto.gov/patents/resources/methods/aiplafall02paper.jsp) on September 21, 2012. Retrieved August 15, 2012.
88. "Debunking the Wayback Machine" (https://web.archive.org/web/20100629050840/http://www.practice.com/2008/12/29/debunking-the-wayback-machine). Archived from the original (http://www.practice.com/2008/12/29/debunking-the-wayback-machine) on June 29, 2010.
89. Bahr, Martin (2002). "The Wayback Machine und Google Cache – eine Verletzung deutschen Urheberrechts?" (http://www.jurpc.de/aufsatz/20020029.htm). *JurPC* (in German): 9. doi:10.7328/jurpcb/20021719 (https://doi.org/10.7328%2Fjurpcb%2F20021719). Archived (https://web.archive.org/web/20090823062928/http://www.jurpc.de/aufsatz/20020029.htm) from the original on August 23, 2009.
90. "Internet Archive FAQ" (https://www.archive.org/about/faqs.php). Archived (https://web.archive.org/web/20140417122600/https://archive.org/about/faqs.php) from the original on April 17, 2014. Retrieved April 16, 2014.
91. Bowman, Lisa M (September 24, 2002). "Net archive silences Scientology critic" (https://web.archive.org/web/20120515210932/http://news.cnet.com/2100-1023-959236.html). CNET News. Archived from the original (http://news.cnet.com/2100-1023-959236.html) on May 15, 2012. Retrieved January 4, 2007.
92. Jeff (September 23, 2002). "exclusions from the Wayback Machine" (https://www.archive.org/iathreads/post-view.php?id=778) (Blog). *Wayback Machine Forum*. Internet Archive. Archived (https://web.archive.org/web/20070211155138/http://www.archive.org/iathreads/post-view.php?id=778) from the original on February 11, 2007. Retrieved January 4, 2007. *Author and Date indicate initiation of forum thread*.
93. Miller, Ernest. "Sherman, Set the Wayback Machine for Scientology" (https://web.archive.org/web/20121116072812/http://lawmeme.research.yale.edu/modules.php?name=News&file=article&sid=350). *LawMeme*. Yale Law School. Archived from the original (http://research.yale.edu/lawmeme/modules.php?name=News&file=article&sid=350) (Blog) on November 16, 2012. Retrieved January 4, 2007.
94. Dye, Jessica (2005). "Website Sued for Controversial Trip into Internet Past". *EContent*. 28. **11**: 8–9.
95. Bangeman, Eric (August 31, 2006). "Internet Archive Settles Suit Over Wayback Machine" (https://arstechnica.com/news.ars/post/20060831-7634.html). *Ars Technica*. Archived (https://web.archive.org/web/20071105205430/http://arstechnica.com/news.ars/post/20060831-7634.html) from the original on November 5, 2007. Retrieved November 29, 2007.
96. *Internet Archive v. Shell*, 505 F.Supp.2d 755 at justia.com (http://dockets.justia.com/docket/colorado/codce/1:2006cv01726/98194), 1:2006cv01726 (Colorado District Court August 31, 2006) ("'April 25, 2007 Settlement agreement announced.' Filing 65, 2007-04-30: '...therefore ORDERED that this matter shall be DISMISSED WITH PREJUDICE...'").
97. Babcock, Lewis T. (February 13, 2007). "*Internet Archive v. Shell* Civil Action No. 06cv01726LTBCBS" (http://blog.ericgoldman.org/archives/waybackshell.pdf) (PDF). Archived (https://web.archive.org/web/20140125143809/http://blog.ericgoldman.org/archives/waybackshell.pdf) (PDF) from the original on January 25, 2014. Retrieved March 25, 2015. "1) Internet Archive's motion to dismiss Shell's counterclaim for conversion and civil theft (Second Cause of Action) is GRANTED, 2) Internet Archive's motion to dismiss Shell's counterclaim for breach of contract (Third Cause of Action) is DENIED; 3) Internet Archive's motion to dismiss Shell's counterclaim for Racketeering under RICO and COCCA (Fourth Cause of Action) is GRANTED."

98. Claburn, Thomas (March 16, 2007). "Colorado Woman Sues To Hold Web Crawlers To Contracts" (http://www.informationweek.com/colorado-woman-sues-to-hold-web-crawlers-to-contracts/d/d-id/1053075). New York, New York, US: InformationWeek, UBM Tech, UBM LLC. Archived (https://web.archive.org/web/20140904170742/http://www.informationweek.com/colorado-woman-sues-to-hold-web-crawlers-to-contracts/d/d-id/1053075) from the original on September 4, 2014. Retrieved March 25, 2015. "Computers can enter into contracts on behalf of people. The Uniform Electronic Transactions Act (UETA) says that a 'contract may be formed by the interaction of electronic agents of the parties, even if no individual was aware of or reviewed the electronic agents' actions or the resulting terms and agreements.' "
99. Samson, Martin H. (2007). "Internet Archive v. Suzanne Shell" (http://www.internetlibrary.com/cases/lib_case456.cfm). Internet Library of Law and Court Decisions. Archived (https://web.archive.org/web/20140803002113/http://www.internetlibrary.com/cases/lib_case456.cfm) from the original on August 3, 2014. Retrieved March 25, 2015. "More importantly, held the court, Internet Archive's mere copying of Shell's site, and display thereof in its database, did not constitute the requisite exercise of dominion and control over defendant's property. Importantly, noted the court, the defendant at all times owned and operated her own site. Said the Court: 'Shell has failed to allege facts showing that Internet Archive exercised dominion or control over her website, since Shell's complaint states explicitly that she continued to own and operate the website while it was archived on the Wayback machine. Shell identifies no authority supporting the notion that copying documents is by itself enough of a deprivation of use to support conversion. Conversely, numerous circuits have determined that it is not.' "
100. brewster (April 25, 2007). "Internet Archive and Suzanne Shell Settle Lawsuit" (https://archive.org/post/119669/lawsuit-settled). *Internet Archive*. Denver, CO, US. Archived (https://web.archive.org/web/20101205074057/https://archive.org/post/119669/lawsuit-settled) from the original on December 5, 2010. Retrieved March 25, 2015. "Both parties sincerely regret any turmoil that the lawsuit may have caused for the other. Neither Internet Archive nor Ms. Shell condones any conduct which may have caused harm to either party arising out of the public attention to this lawsuit. The parties have not engaged in such conduct and request that the public response to the amicable resolution of this litigation be consistent with their wishes that no further harm or turmoil be caused to either party."
101. Stobbe, Richard (December 5, 2014). "Copyright Implications Of A "Right To Be Forgotten"? Or How To Take-Down The Internet Archive" (http://www.mondaq.com/canada/x/358282/Trade%20Secrets/Copyright%20Implications%20of%20a%20Right%20to%20be%20Forgotten%20Or%20How%20to%20TakeDown%20the%20Internet%20Archive). *Mondaq*. Archived (https://web.archive.org/web/20181118095257/http://www.mondaq.com/canada/x/358282/Trade%20Secrets/Copyright%20Implications%20of%20a%20Right%20to%20be%20Forgotten%20Or%20How%20to%20TakeDown%20the%20Internet%20Archive) from the original on November 18, 2018. Retrieved March 8, 2019.
102. McVeigh, Glennys (October 16, 2014). Philpott, James; Weissman, Adam; Bucholz, Ren; Kettles, Brent; Pearl, Aaron (eds.). "Davydiuk v. Internet Archive Canada, 2014 FC 944" (http://canlii.ca/t/gf3d7). *CanLII*. Federation of Law Societies of Canada. Archived (https://web.archive.org/web/20201218143737/https://www.canlii.org/en/ca/fct/doc/2014/2014fc944/2014fc944.html) from the original on December 18, 2020. Retrieved March 8, 2019.
103. Southcott, Richard F. (November 30, 2016). Philpott, John; Alton, Alex; Bucholz, Ren (eds.). "Davydiuk v. Internet Archive Canada and Internet Archive, 2016 FC 1313 (CanLII)" (https://www.canlii.org/en/ca/fct/doc/2016/2016fc1313/2016fc1313.html). *CanLII*. Ottawa, Ontario: Federation of Law Societies of Canada. Archived (https://web.archive.org/web/20190629215557/https://www.canlii.org/en/ca/fct/doc/2016/2016fc1313/2016fc1313.html) from the original on June 29, 2019. Retrieved March 8, 2019.
104. Cox, Joseph (May 22, 2018). "The Wayback Machine Is Deleting Evidence of Malware Sold to Stalkers" (https://www.vice.com/en/article/nekzzq/wayback-machine-deleting-evidence-flexispy). *Vice*. Archived (https://web.archive.org/web/20220124045413/https://www.vice.com/en/article/nekzzq/wayback-machine-deleting-evidence-flexispy) from the original on January 24, 2022. Retrieved January 24, 2022.
105. Conger, Kate (December 8, 2016). "Backing up the history of the internet in Canada to save it from Trump" (https://techcrunch.com/2016/12/08/backing-up-the-history-of-the-internet-in-canada-to-save-it-from-trump/). *TechCrunch*. Archived (https://web.archive.org/web/20161227100714/https://techcrunch.com/2016/12/08/backing-up-the-history-of-the-internet-in-canada-to-save-it-from-trump/) from the original on December 27, 2016. Retrieved May 14, 2017.
106. "Where to find what's disappeared online, and a whole lot more: the Internet Archive" (https://www.pri.org/stories/2017-02-23/where-find-whats-disappeared-online-and-whole-lot-more-internet-archive). Public Radio International. February 23, 2017. Archived (https://web.archive.org/web/20170328092846/https://www.pri.org/stories/2017-02-23/where-find-whats-disappeared-online-and-whole-lot-more-internet-archive) from the original on March 28, 2017. Retrieved May 14, 2017.
107. "Online Censorship In China" (https://en.greatfire.org/archive.org). *GreatFire*. October 22, 2023. Retrieved October 22, 2023.

108. Chirgwin, Richard (September 1, 2015). "There's no Wayback in Russia: Putin blocks Archive.org" (https://www.theregister.co.uk/2015/09/01/russias_putin_blocks_archiveorg/). *The Register*. Archived (https://web.archive.org/web/20161007165431/http://www.theregister.co.uk/2015/09/01/russias_putin_blocks_archiveorg) from the original on October 7, 2016. Retrieved May 14, 2017.
109. "Russia won't go Wayback, blocks the Internet Archive" (https://www.digitaltrends.com/web/russia-censors-wayback-machines-485-billion-archived-web-pages/). Digital Trends. June 26, 2015. Archived (https://web.archive.org/web/20160417125727/http://www.digitaltrends.com/web/russia-censors-wayback-machines-485-billion-archived-web-pages/) from the original on April 17, 2016. Retrieved May 14, 2017.
110. "В России разблокирован крупнейший интернет-архив" (https://rg.ru/2016/04/18/v-rossii-razblokirovan-krupnejshij-internet-arhiv.html). *Российская газета* (in Russian). April 18, 2016. Archived (https://web.archive.org/web/20190405031631/https://rg.ru/2016/04/18/v-rossii-razblokirovan-krupnejshij-internet-arhiv.html) from the original on April 5, 2019. Retrieved October 18, 2020.
111. The VirusTotal Team (March 25, 2015). "207.241.226.190 IP address information" (https://www.virustotal.com/en/ip-address/207.241.226.190/information/). *virustotal.com*. Dublin, Ireland: VirusTotal. Archived (https://web.archive.org/web/20140714232311/https://www.virustotal.com/en/ip-address/207.241.226.190/information/) from the original on July 14, 2014. Retrieved March 25, 2015. "2015-03-25: Latest URLs hosted in this IP address detected by at least one URL scanner or malicious URL dataset. ... 2/62 2015-03-25 16:14:12 [complete URL redacted]/Renegotiating_TLS.pdf ... 1/62 2015-03-25 04:46:34 [complete URL redacted]/CBLightSetup.exe"
112. "Safe Browsing Diagnostic page for archive.org" (https://www.google.com/safebrowsing/diagnostic?site=archive.org). *google.com/safebrowsing*. Mountain View, CA, US. March 25, 2015. Archived (https://web.archive.org/web/20150406101743/http://www.google.com/safebrowsing/diagnostic?site=archive.org) from the original on April 6, 2015. Retrieved March 25, 2015. "2015-03-25: Part of this site was listed for suspicious activity 138 time(s) over the past 90 days. ... What happened when Google visited this site? ... Of the 42410 pages we tested on the site over the past 90 days, 450 page(s) resulted in malicious software being downloaded and installed without user consent. The last time Google visited this site was on 2015-03-25, and the last time suspicious content was found on this site was on 2015-03-25. ... Malicious software includes 169 trojan(s), 126 virus, 43 backdoor(s)."
113. "Help Us Keep the Archive Free, Accessible, and Reader Private | Internet Archive Blogs" (https://blog.archive.org/2016/11/29/help-us-keep-the-archive-free-accessible-and-private/). November 29, 2016. Archived (https://web.archive.org/web/20170521080543/http://blog.archive.org/2016/11/29/help-us-keep-the-archive-free-accessible-and-private/) from the original on May 21, 2017. Retrieved May 14, 2017.
114. Sakr, Sharif (November 7, 2013). "Wayback Machine web archive survives destructive fire but needs help to recover" (https://web.archive.org/web/20201109025136/https://www.engadget.com/2013-11-07-wayback-machine-internet-archive-fire.html). Archived from the original (https://www.engadget.com/2013-11-07-wayback-machine-internet-archive-fire.html) on November 9, 2020. Retrieved April 3, 2022.
115. "Internet Archive: Proposed Changes To DMCA Would Make Us "Censor The Web" " (https://consumerist.com/2016/06/07/internet-archive-proposed-changes-to-dmca-would-make-us-censor-the-web/). Consumerist. June 7, 2016. Archived (https://web.archive.org/web/20161111161239/https://consumerist.com/2016/06/07/internet-archive-proposed-changes-to-dmca-would-make-us-censor-the-web/) from the original on November 11, 2016. Retrieved May 14, 2017.
116. Herb, Ulrich (December 6, 2016). "Die Trump-Angst grassiert" (https://www.heise.de/tp/features/Die-Trump-Angst-grassiert-3549579.html) [The Trump fear is rampant] (in German). heise online. Archived (https://web.archive.org/web/20161207164413/https://www.heise.de/tp/features/Die-Trump-Angst-grassiert-3549579.html) from the original on December 7, 2016. Retrieved May 14, 2017. "Die Betreiber des Archives fürchten neben Zensur und Manipulation der digitalen Aufzeichnungen demnach auch die Überwachung der Archive-Nutzer. [The operators of the archive fear not only censorship and manipulation of the digital recordings, but also the surveillance of the archive users]."
117. LaFrance, Adrienne (October 14, 2015). "The Internet's Dark Ages" (https://www.theatlantic.com/technology/archive/2015/10/raiders-of-the-lost-web/409210/). *The Atlantic*. Archived (https://web.archive.org/web/20170507173716/https://www.theatlantic.com/technology/archive/2015/10/raiders-of-the-lost-web/409210/) from the original on May 7, 2017. Retrieved May 14, 2017.
118. "The Entire Internet Will Be Archived In Canada to Protect It From Trump" (https://motherboard.vice.com/en_us/article/the-entire-internet-will-be-archived-in-canada-to-protect-it-from-trump). Motherboard. November 29, 2016. Archived (https://web.archive.org/web/20170516221604/https://motherboard.vice.com/en_us/article/the-entire-internet-will-be-archived-in-canada-to-protect-it-from-trump) from the original on May 16, 2017. Retrieved May 14, 2017.
119. LaFrance, Adrienne (June 3, 2016). "The Human Fear of Total Knowledge" (https://www.theatlantic.com/technology/archive/2016/06/knowledge-compendia/485507/). *The Atlantic*. Archived (https://web.archive.org/web/20161202040113/http://www.theatlantic.com/technology/archive/2016/06/knowledge-compendia/485507/) from the original on December 2, 2016. Retrieved May 14, 2017.

120. Shaikh, Roshan (October 10, 2024). "Internet Archive hacked and 31 million user accounts leaked — hacking group 'SN_Blackmeta' claims responsibility" (https://www.tomshardware.com/tech-industry/cyber-security/internet-archive-hacked-and-31-million-user-accounts-leaked-hacking-group-sn-blackmeta-claims-responsibility). *Tom's Hardware*. Retrieved December 30, 2024.
121. Goodin, Dan (October 9, 2024). "Archive.org, a repository of the history of the Internet, has a data breach" (https://arstechnica.com/information-technology/2024/10/archive-org-a-repository-storing-the-entire-history-of-the-internet-has-been-hacked/). *Ars Technica*. Archived (https://archive.today/20241010115326/https://arstechnica.com/information-technology/2024/10/archive-org-a-repository-storing-the-entire-history-of-the-internet-has-been-hacked/) from the original on October 10, 2024. Retrieved October 10, 2024.
122. Warren, Tom (October 14, 2024). "The Internet Archive is back as a read-only service after cyberattacks" (https://www.theverge.com/2024/10/14/24269741/internet-archive-online-read-only-data-breach-outage). *The Verge*. Retrieved October 14, 2024.
123. "Wayback Machine Homepage" (https://web.archive.org/web/20241024041605/https://web.archive.org/). October 24, 2024. Archived from the original (https://web.archive.org/) on October 24, 2024.

External links

- Official website (https://web.archive.org)
- *Internet history is fragile. This archive is making sure it doesn't disappear* (https://www.youtube.com/watch?v=K8I28erYFLc). *PBS Newshour*. Archived (https://web.archive.org/web/20220106023827/https://www.youtube.com/watch?v=K8I28erYFLc) from the original on January 6, 2022. Retrieved November 5, 2024.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Wayback_Machine&oldid=1266222054"